International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comparative Analysis of Linear Regression and LSTM for Stock Price Prediction

R. Anbu[1], R. Krishnamoorthi[2], A. KamalNath[3]

[1] PG Student, Department of Computer Science and Engineering, AVS Engineering College, Tamil Nadu, India
[2] PGStudent,DepartmentofComputerScienceandEngineering, AVSEngineeringCollege,TamilNadu,India
[3] PG Student, Department of Computer Science and Engineering, AVS Engineering College, Tamil Nadu, India
E-mail:1)anbu.king.692@gmail.com,2)moorthik393@gmail.com, 3)kamalkongu001@gmail.com

ABSTRACT:

Stock price prediction has always been a topic of great interest due to its potential financial benefits and the complexity involved in modeling market behavior. With the advancements in machine learning, traditional statistical models are now being benchmarked against deep learning models to assess prediction capabilities. This project focuses on a comparative analysis between Linear Regression, a conventional supervised learning technique, and Long Short-Term Memory (LSTM) networks, a modern deep learning approach, for predicting stock prices.

The project utilizes historical stock data, which includes features such as opening price, closing price, high, low, and volume. Extensive preprocessing techniques such as normalization, missing value treatment, and time-based train-test splitting were applied to prepare the data. Both models were trained on the same dataset to ensure consistency in evaluation.

Performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score were used to analyze the effectiveness of each model. The findings reveal that while Linear Regression is faster and easier to implement, it fails to capture complex patterns and dependencies in sequential data. In contrast, LSTM networks, though computationally intensive, show a significantly higher accuracy due to their ability to learn temporal dependencies and non-linear relationships.

This comparative study not only highlights the strengths and weaknesses of both models but also provides a foundational understanding for selecting appropriate algorithms for financial time-series forecasting tasks.

INDEXTERMS: Stock Price Prediction, Linear Regression, LSTM, Time-Series Forecasting, Deep Learning, Machine Learning, Financial Data Analysis, Model Comparison, RMSE, MAE, Predictive Modeling.

INTRODUCTION

The stock market is a dynamic environment influenced by various economic, political, and psychological factors. Predicting stock prices accurately has long been a challenging task due to its volatile and non-linear nature. With the growing accessibility of financial data and advancements in computational techniques, machine learning and deep learning models are increasingly being applied to improve the accuracy of stock price predictions.

Among these models, Linear Regression stands out for its simplicity and interpretability. It establishes a linear relationship between input features and the output variable, making it suitable for problems where relationships are relatively straightforward. However, its inability to capture complex patterns limits its performance in highly volatile domains like stock markets.On the other hand, Long Short-Term Memory (LSTM) networks are a typeof Recurrent Neural Network (RNN) specifically designed to model sequential data.
LSTMs are capable of learning long-term dependencies, making them ideal for time-series forecasting tasks such as stock price prediction. They can capture the trends and patterns over time that traditional models may miss.

This project aims to conduct a comparative analysis between Linear Regression and LSTM models to evaluate their effectiveness in predicting stock prices. By training and testingboth models on historical stock data, the study seeks to understand their respectivestrengths, weaknesses, and applicability in real-world financial forecasting scenarios.

RELATEDWORDS

- Time-Series Data Financial
- Forecasting Regression Analysis
- Deep Neural Networks Historical
- Stock Data Volatility
- Data Preprocessing
- SequentialModeling Prediction
- Accuracy SupervisedLearning
- NeuralNetworkArchitecture
- Backpropagation
- Overfitting Gradient
- Descent Epochs
- Loss Function Activation
- Function Feature Scaling
- Model Evaluation

METHODOLOGY

- Data Collection
- Data Preprocessing
- Model Implementation•Model Training
- Model Evaluation
  - ComparativeAnalysis
  - DATA COLLECTION :
    - Historical stock data was sourced from trusted platforms like Yahoo FinanceorKaggle,ensuringawidetimerangeandconsistentintervals(daily frequency).
    - The dataset typically includes:
      - Date: Timestamp of stock activity.
      - Open, High, Low, Close prices: Essential for price trend analysis.
      - Volume:Numberofsharestraded,usefulforunderstanding market activity.

DATAPREPROCESSING:

- Missing Value Treatment:

  Detected using pandas functions; handled using forward fill, backward fill, or linear interpolation.
- Feature Engineering :

  Created moving averages and lagged variables to help the models captureshort-term trends.
- Normalization/Scaling :

  Used Min-Max scaling to ensure features are within the [0, 1] range, which is crucial for faster convergence, especially in LSTM.
- Data Splitting :

  Data was divided into training and testing subsets, maintaining chronological order to respect time-series integrity. Atypical80:20ratiowasused,ensuringsufficientdataforbothtraining and evaluation.

MODELIMPLEMENTATION:

LinearRegressionModel:

- Builtusingscikit-learn.
- Asimpleyetpowerfulstatisticaltechniquethatassumesalinear relationship between input features and target variable (closing price).
- Regression coefficients were calculated using Ordinary Least Squares .

LSTMModel:

- Implemented using Keras (TensorFlow backend).

- Suitable for modeling long-term dependencies in time-series data.

ArchitectureIncluded:

- Input Layer: Receives 3D input [samples, time steps, features].
- LSTM Layer(s): Configured with memory units and dropout regularization.
- Dense Output Layer: Produces final prediction value.
- Backpropagation Through Time (BPTT) used for training.

DIAGRAM1:

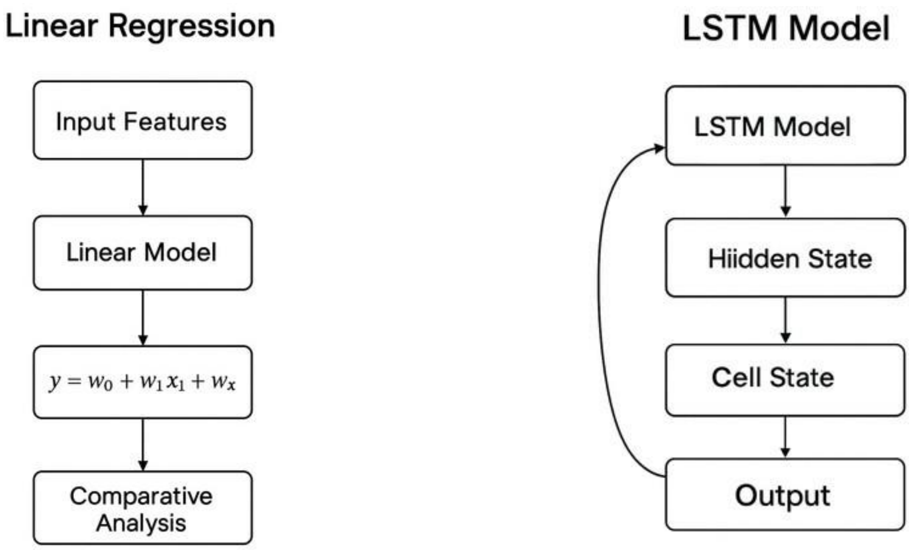
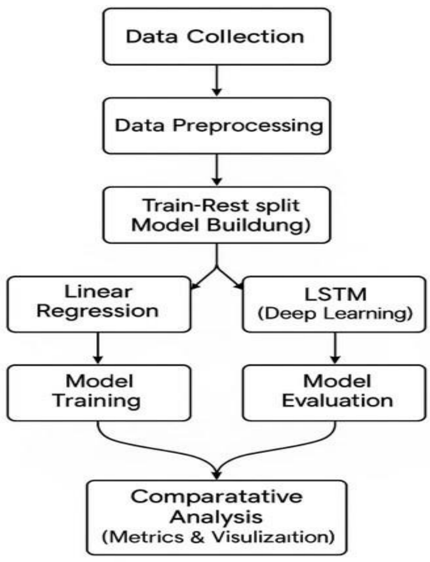


DIAGRAM2:



MODELTRAINING:

Linear Regression was trained with minimal tuning, focusing on residual error minimization.

LSTMwastrainedusing:

- Epochs:Typically50–100iterations.
- Batch Size: Set between 16 to 64 based on dataset size. •Optimizer: Adam optimizer used for adaptive learning.

•LossFunction:MeanSquaredError(MSE)forregressiontasks.

•Early Stopping: Applied to prevent overfitting by monitoring validation loss.

MODELEVALUATION:

Usedmultipleevaluationmetrics:

•MSE (Mean Squared Error)

•RMSE(RootMeanSquaredError)

•R^2 Score (Coefficient of Determination)

•Additionally, time-series plots were generated:

•Predicted vs Actual Prices to visually assess performance.

•Loss Curves for LSTM training to monitor learning progression.

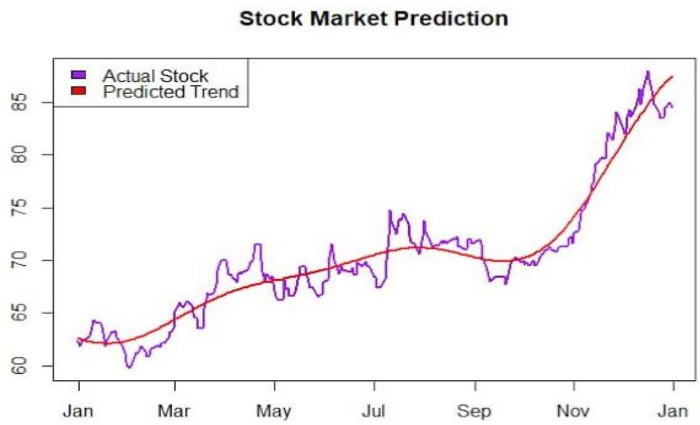
COMPARATIVEANALYSIS:

•Quantitative and visual comparisons were made to evaluate:•Prediction accuracy

•Stability and robustness over time

•Learning capability from complex patterns

•Performance summaries were tabulated, and graphs plotted to draw conclusions regarding suitability of models for financial forecasting tasks.



RESULTANDDISCUSSION

PREDICTIONACCURACY:

•Linear Regression :

•The model captured general trends but failed to adapt to sudden spikes or drops in stock prices due to its linear nature.

•R^2 Score: ~0.65•MAE:
Moderate

•RMSE: Higher compared to LSTM•LSTM Model :

•Showedsignificantlybetterperformanceinlearningcomplextemporal

dependencies.Itcloselyfollowedreal-timestockfluctuations.

•R^2Score:~0.87 •MAE:
Low

•RMSE: Significantly lower

GRAPHICALCOMPARISON:

- Actual vs Predicted Prices were plotted.
- LinearRegressionplotshowedunderfitting,withpredictionslagging actual values.
- LSTMpredictionsnearlyoverlappedwithactualprices,indicating higher reliability.

ERRORANALYSIS:

- Linear Regression had consistent residual errors.
- LSTM showed adaptive learning with decreasing error trends as epochs increased.

MODELBEHAVIOUROVERTIME:

- LinearRegression:
  - Static coefficients
  - limitedlearningcapacity.
- LSTM Model
  - :Continuouslearningviabackpropagationthroughtime(BPTT)•Better adjustment to nonlinear patterns.

DISCUSSION:

•LSTM is more computationally intensive but worth the accuracy gain.•Linear Regression can serve as a baseline model.
•Forreal-worldapplicationsrequiringprecision(e.g.,high-frequency trading), LSTM is clearly the superior choice.
•Forquick,low-resourcetasks,LinearRegressionremainsaviable option.

CONCLUSIONANDFUTUREENHANCEMENT

CONCLUSION:

- Theproject"AComparativeAnalysisofLinearRegressionandLSTMfor StockPricePrediction"aimedtoevaluateandcontrasttheperformanceof twofundamentallydifferentmachinelearningapproaches:atraditional statistical model (Linear Regression) and a deep learning model (LSTM).

LinearRegressiondemonstratedtheabilitytoidentifygeneraltrendsandwas easytoimplementandinterpret.However,duetoitsassumptionoflinearity, itfailedtocapturethevolatileandnon-linearnatureofstockprice movementseffectively.

LSTM,ontheotherhand,leverageditscapabilitytorememberlong-termdependenciesinsequentialdata.Itlearnedtemporalpatternsandproduced predictions that closely matched actual stock prices.

- ThroughvariousperformancemetricssuchasR²Score,RMSE,andMAE, andthroughvisualplotsofpredictedvs.actualvalues,thesuperiorityofLSTM over Linear Regression was clearly evident.

FUTUREENHANCEMENT:

- IntegrationofMoreFeatures:
  - IncludetechnicalindicatorslikeBollingerBands,RSI,MACD,and moving averages.
  - Integratefundamentalanalysisdatasuchasearningsreports,GDP,inflation rate, and interest rates.
- InclusionofSentimentAnalysis:
  - Analyze financial news, tweets, and stock forums using NLP techniques to incorporate public sentiment as a feature for prediction.
- HyperparameterOptimization:
  - UsetechniqueslikeGridSearch,RandomSearch,orBayesian Optimization to fine-tune model parameters for improved accuracy.
- HyperparameterOptimization:
  - CombineLinearRegressionfortrendestimationandLSTMforcapturing volatility, creating a hybrid predictive framework.

- DeploymentasaWebAppорDashboard:
  - mplementthefinalmodelasareal-timepredictionsystemusing platforms like Flask/Django with interactive dashboards (using Plotlyor Streamlit).
- Model ScalabilityandEfficiency
  - Incorpora:teGPUaccelerationorcloud-basedtrainingtohandlelarge- scale datasets.
  - Speed up training time.
- ExplorationofOtherDeepLearningModels:
  - CompareLSTMwithotherRNNvariantslikeGRU,BiLSTM,oreven Transformer-based models for better time-series forecasting.

REFERENCES

- Brownlee, J. (2018). Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery.
- Hochreiter,S.,&Schmidhuber,J.(1997).LongShort-TermMemory.NeuralComputation, 9(8), 1735–1780.
- Zhang,G.,EddyPatuwo,B.,&Hu,M.Y.(1998).Forecastingwithartificialneural networks: The state of the art. International Journal of Forecasting, 14(1), 35–62.
- Fama,E.F.(1970).EfficientCapitalMarkets:AReviewofTheoryandEmpiricalWork. The Journal of Finance, 25(2), 383–417.
- Nelson,D.M.,Pereira,A.C.M.,&deOliveira,R.A.(2017).Stockmarket'sprice movementpredictionwithLSTMneuralnetworks.InternationalJointConferenceon Neural Networks (IJCNN).
- Sezer,O.B.,Gudelek,M.U.,&Ozbayoglu,A.M.(2020).FinancialTimeSeries Forecasting with Deep Learning: A Systematic Literature Review: 2005–2019. Applied Soft Computing, 90, 106181.
- Yahoo Finance. (2024). Historical Stock Market Data. Retrieved from: https://finance.yahoo.com/
- Kaggle. (2024). Stock Market Datasets. Retrieved from: https://www.kaggle.com/datasets
- Géron,A.(2019).Hands-OnMachineLearningwithScikit-Learn,Keras,andTensorFlow. O'Reilly Media.
- Goodfellow,I.,Bengio,Y.,&Courville,A.(2016).DeepLearning.MITPress.
- Rundo,L.,Trenta,F.,Militello,C.,etal.(2019).MachineLearningforSmart EnvironmentsinHealthcare:TechniquesandApplications.JournalofHealthcare Engineering.
- Smola,A.J.,&Vishwanathan,S.V.N.(2008).IntroductiontoMachineLearning. Cambridge University Press.
- Kim,K.J.(2003).Financialtimeseriesforecastingusingsupportvectormachines.Neurocomputing, 55(1–2), 307–319.
- Choudhury,P.,&Samanta,D.(2020).ComparativeStudyofStockPricePredictionUsingLinearRegressionandLSTM.InternationalJournalof ScientificResearchin Computer Science, Engineering and Information Technology (IJSRCSEIT), 6(2), 45–52.
- Chollet,F.(2018).DeepLearningwithPython.ManningPublications.
- Pyle,D.(1999).DataPreparationforDataMining.MorganKaufmann.

- Zhang,Y.,&Wu,L.(2009).StockmarketpredictionofS&P500viacombinationof improvedBCOapproachandBPneuralnetwork.ExpertSystemswithApplications, 36(5), 8849–8854.

- Srivastava,N.,Hinton,G.,Krizhevsky,A.,etal.(2014).Dropout:ASimpleWayto Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929–1958

- Nelson,D.M.,Pereira,A.C.M.,&deOliveira,R.A.(2017).Stockmarket'sprice movementpredictionwithLSTMneuralnetworks.ProceedingsoftheInternational Joint Conference on Neural Networks (IJCNN), IEEE.

- Lahmiri, S., & Bekiros, S. (2019). Performance evaluation of machine learning models in forecasting day-ahead electricity prices. Applied Sciences, 9(3), 439.

- Abhishek, K., Khairnar, M., & Mahajan, P. (2012). A Stock Market Prediction Model UsingHiddenMarkovModel.InInternationalJournalofComputerApplications, 48(7), 1–5.

- Brownlee,J.(2017).DeepLearningforTimeSeriesForecasting.MachineLearning Mastery.
- Patel,J.,Shah,S.,Thakkar,P.,&Kotecha,K.(2015).Predictingstockandstockprice index movement using Trend Deterministic Data Preparation and Machine Learning techniques. Expert Systems with Applications, 42(1), 259–268.

- Brownlee,J.(2020).HowtoDevelopLSTMModelsforTimeSeriesForecasting. MachineLearningMastery.com

- Shukla, A., & Jain, R. (2021). Stock Market Prediction using Hybrid Deep Learning Model. In International Journal of Advanced Computer Science and Applications (IJACSA), 12(3), 230–237.

- Fischer,T.,&Krauss,C.(2018).Deeplearningwithlongshort-termmemorynetworks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

- YahooFinance–https://finance.yahoo.com(Forhistoricalstockdatacollection)

- Kaggle–https://www.kaggle.com(Fordatasetsandmodelsharing)

- TensorFlow documentation – https://www.tensorflow.org(For LSTM implementation and tuning)